

데이터 마이닝 응용 프로젝트 중간발표

지속가능한 전력 생산

전력 수요, 태양열 발전량, 풍력 발전량 예측

TEAM. 뽕뽕이들

1. 데이터 설명

| 데이터 세부 설명

범위 : 전국

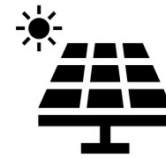
기간 : 2020 ~ 2022년 (3개년)



기상



풍력 발전량



태양광 발전량

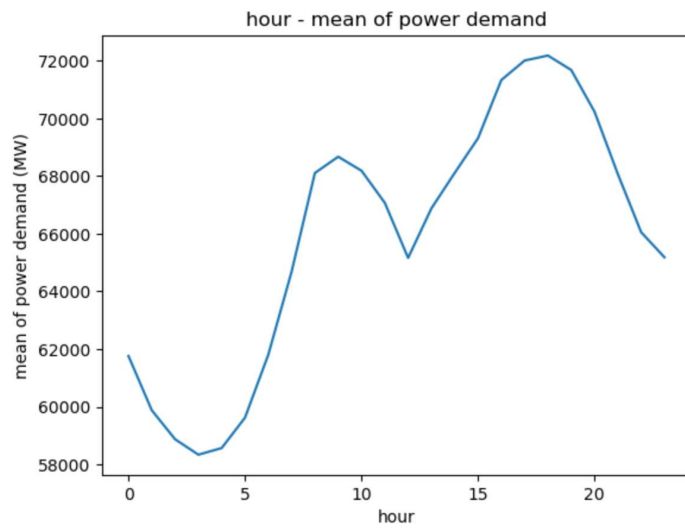


전력

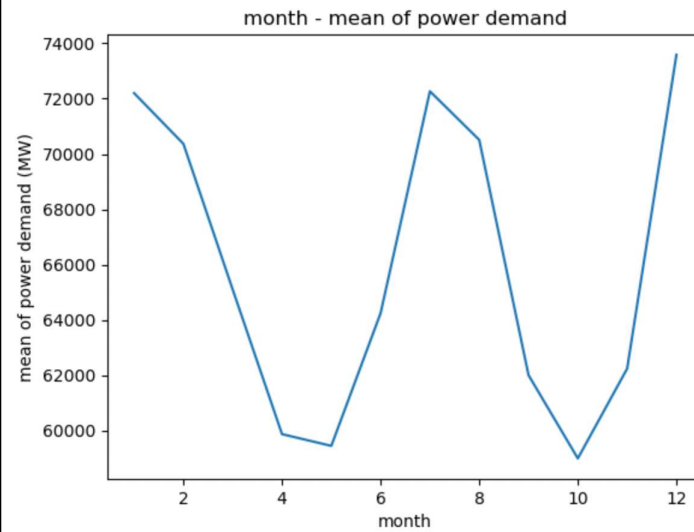
1. 데이터 설명

| 데이터 세부 설명

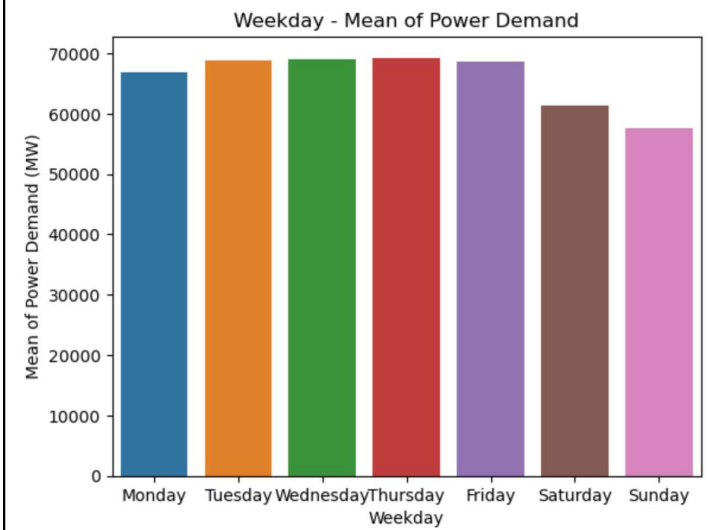
시간대별 전력 수요량



월별 전력 수요량



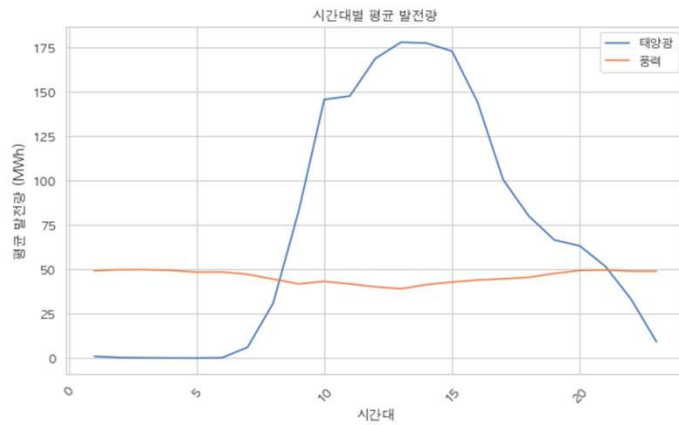
요일별 전력 수요량



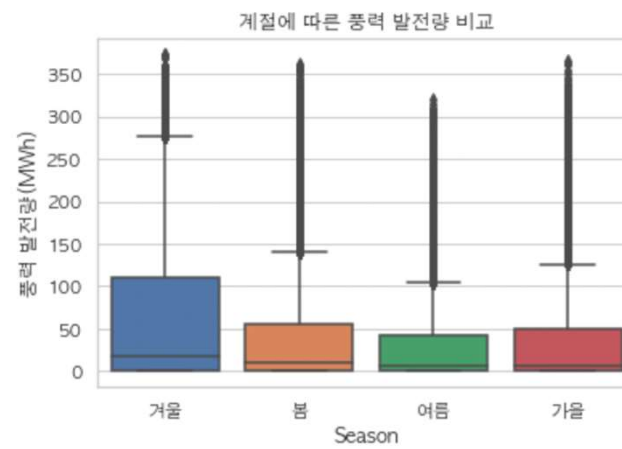
1. 데이터 설명

| 데이터 세부 설명

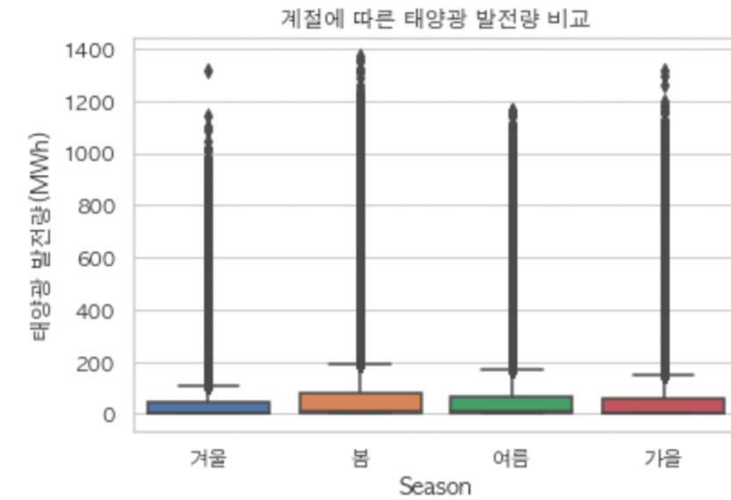
시간대별 발전량



계절별 풍력 발전량



계절별 태양광 발전량



2. 데이터 전처리

관측 : 501개소



+

구역 : 16 지점



+

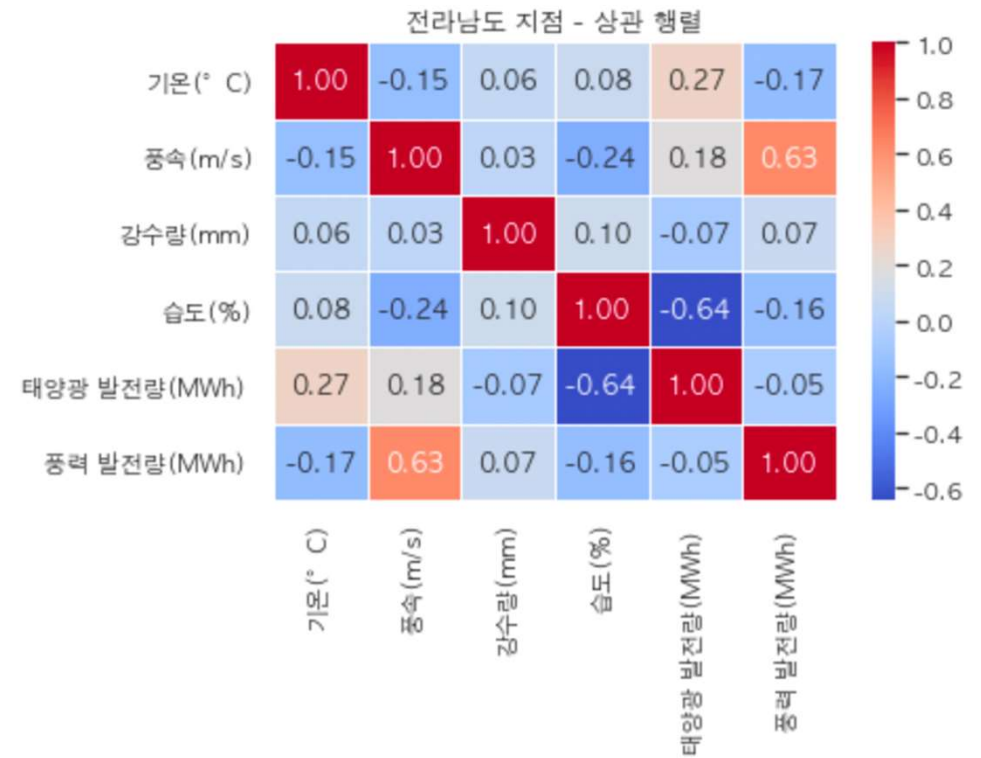
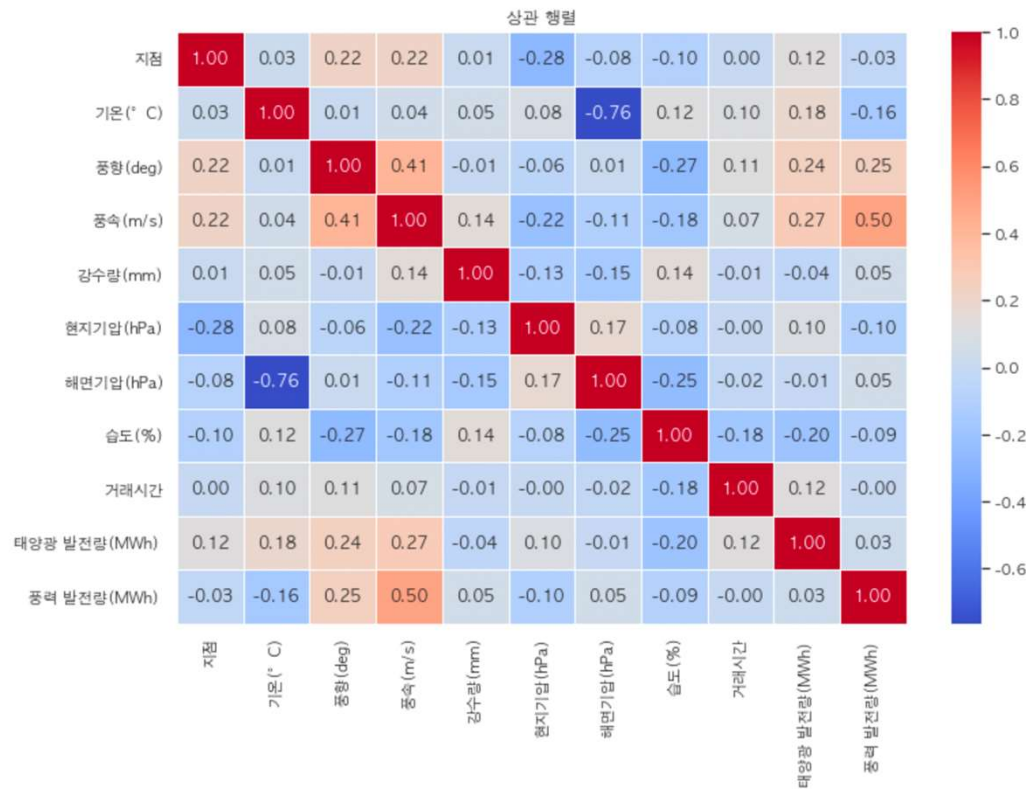
구역 : 16 지점



구분	지점	일시	기온	풍향	풍속	강수량	현지기압	해면기압	습도	도시명	태양광 발전량	풍력발전량
데이터 타입	int	datetime	float	float	float	float	float	float	float	string	float	float

2. 데이터 전처리

| 데이터 상관관계 (피어슨 상관계수)



3. 모델 설명

| 단일 모델_풍력&태양광

단일 모델을 위한 데이터 셋

구분	구성
기상요소	습도, 풍속, 강우량, 풍향
시간 데이터	년, 월, 일
위치 데이터	도시명
예측 목표값	태양광 발전량, 풍력 발전량

단일 모델을 위한 데이터 전처리

요소명	전처리 방법
습도, 풍속, 강우량, 풍향	표준화
도시명	One-Hot Encoding을 통해 벡터화

3. 모델 설명

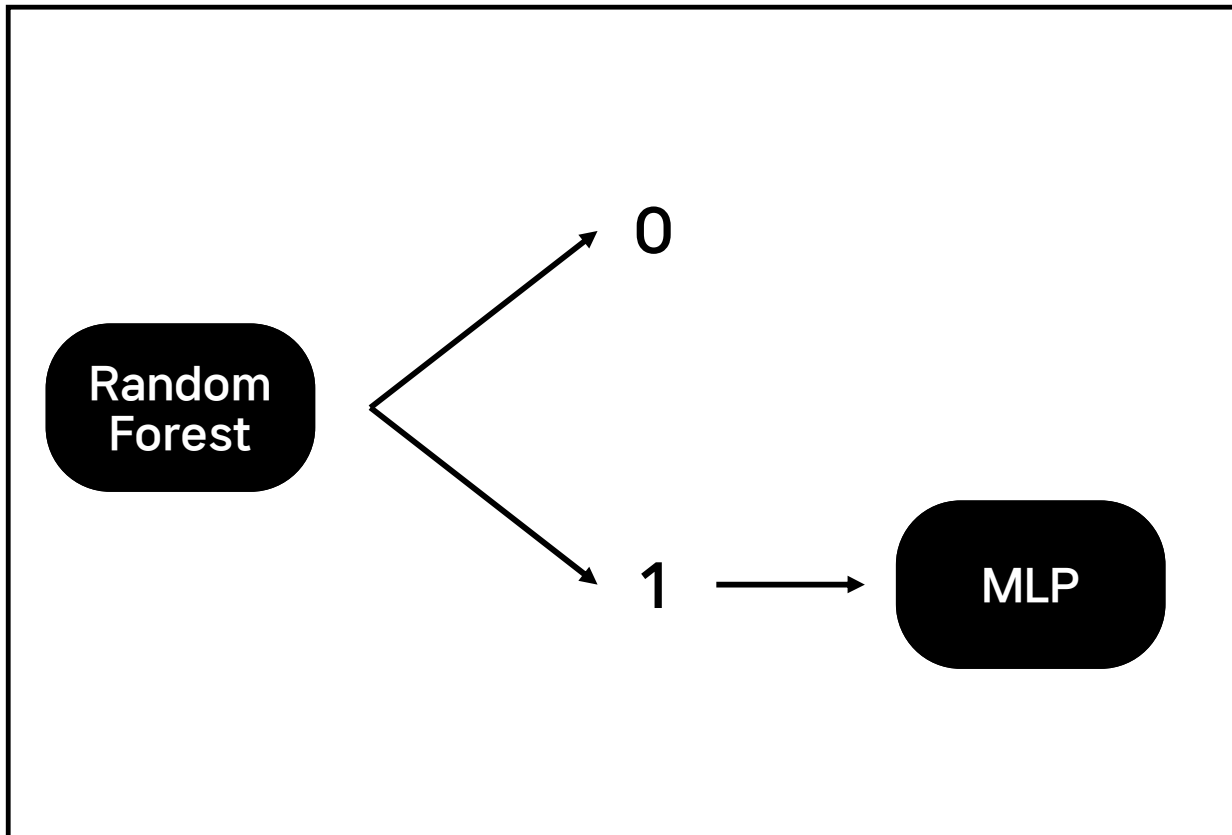
| 단일 모델_풍력&태양광

모델명	RMSE		parameter
	풍력 에너지	태양에너지	
Decision Tree	47.57	77.42	max_depth=8
Multiple Linear Regression	52.24	108.63	
Random Forest	41.48	54.86	max_depth=8
Multi-Layer Perceptron	41.08	49.69	max_iter=500

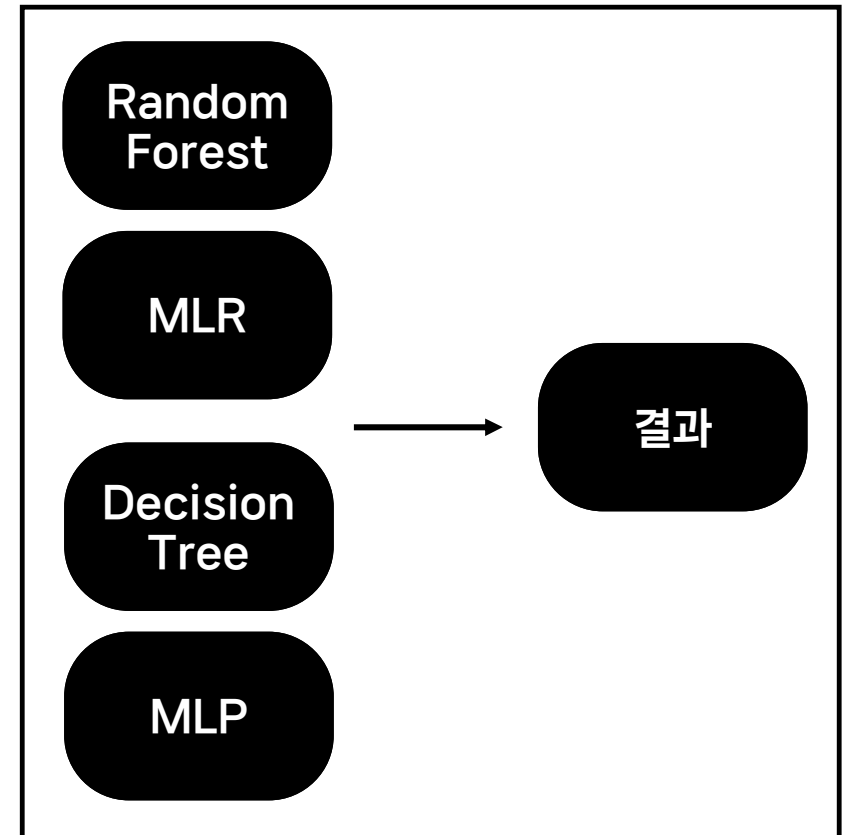
3. 모델 설명

| 모델 결합

모델 예시1



모델 예시2



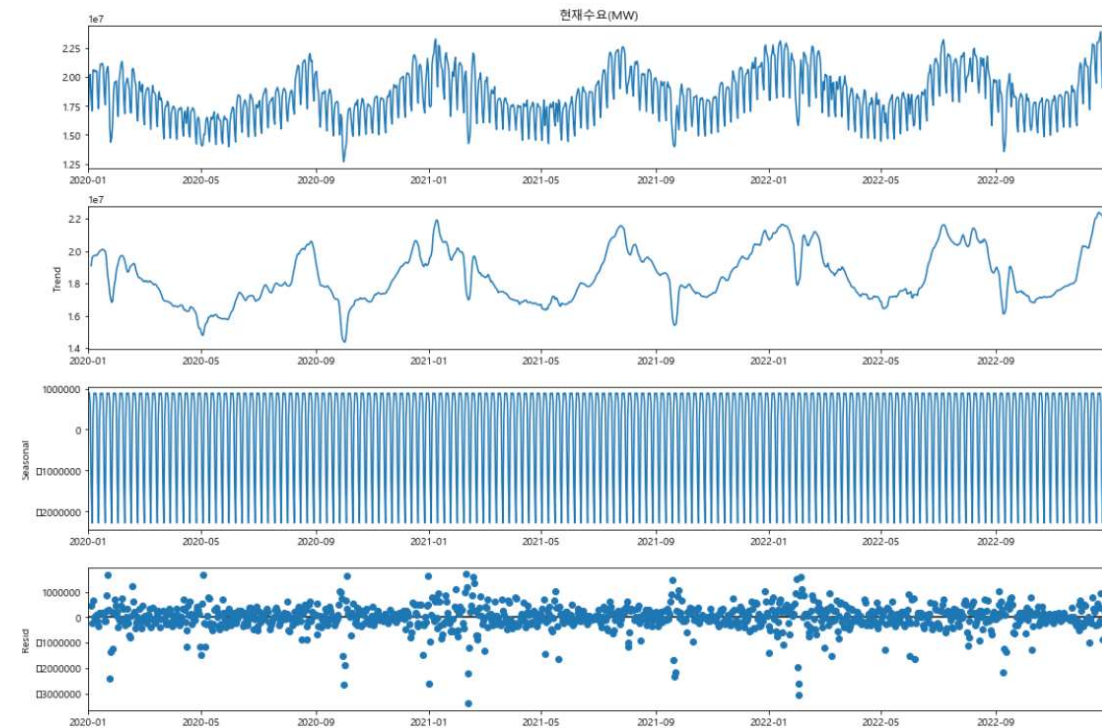
3. 모델 설명

1. 시계열분석의 목적

과거 데이터 바탕으로 미래 '전력 수요' 예측

2. 시계열분석 - ARIMA 선택 이유

- 1) 단기 예측
- 2) 계절적 변동 요인
- 3) 데이터 수(Sample) > 50



시계열 분해 결과

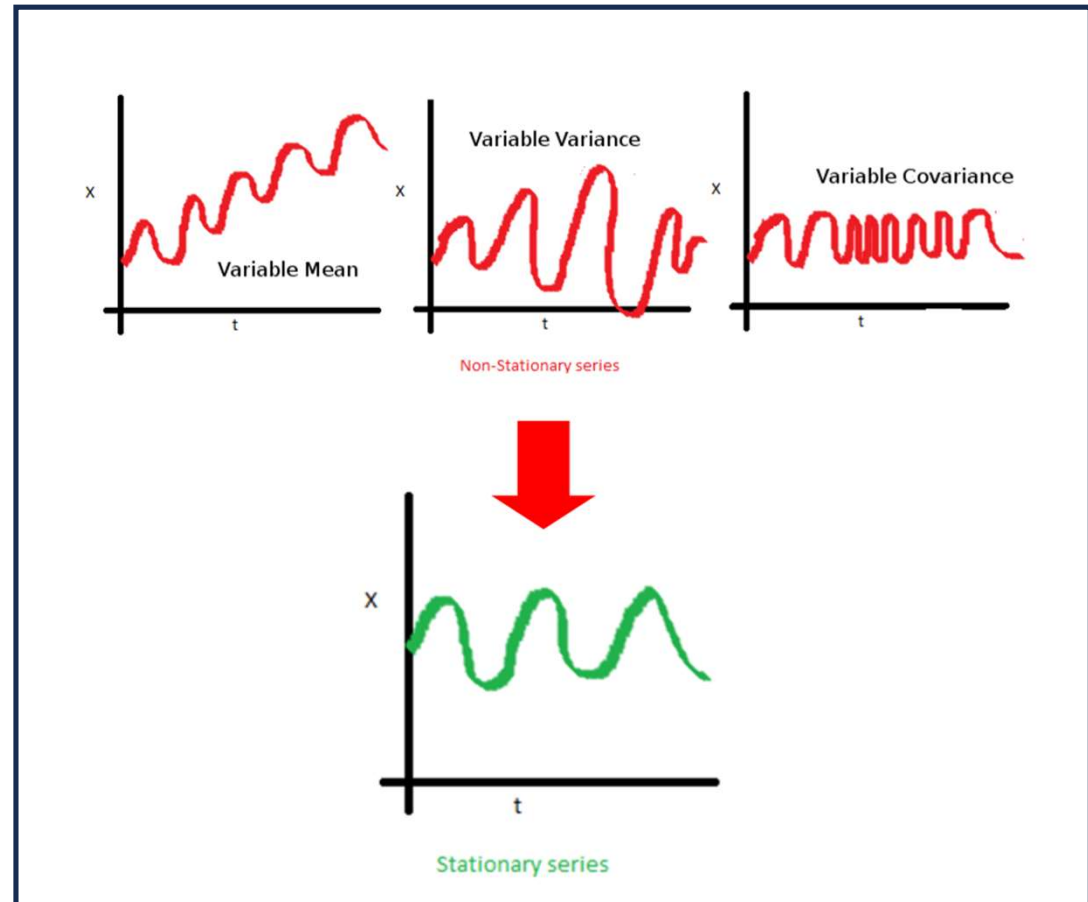
- 추세 존재 X
- 계절성 · 주기성 존재 O

3. 모델 설명

3. 시계열분석 모델 적용 조건

★ 정상성 만족

- 시간에 따른 평균 일정
- 시간에 따른 분산 일정
- 시간에 따른 공분산 일정



3. 모델 설명

=> 정상성 검증

KPSS Statistic: 0.767720394093419

p-value: 0.01

Lags Used: 52

Critical Values: {'10%': 0.347, '5%': 0.463,

검증결과: 비정상(non-stationary) 시계열 데이터입니다.

ADF Statistic: -8.881919853667899

p-value: 1.3109363183008864e-14

Critical Values: {'1%': -3.4310999685396

Best number of lags (AIC): 8722

검증결과: 정상(stationary) 시계열 데이터입니다.

1차 차분



KPSS Statistic after differencing: 0.00

p-value after differencing: 0.1

Lags Used after differencing: 40

Critical Values after differencing: {'1

검증결과: 정상(stationary) 시계열 데이터입니다.

ADF Statistic: -15.307786341211989

p-value: 4.226749935307493e-28

Critical Values: {'1%': -3.431100054560

Best number of lags (AIC): 8721

검증결과: 정상(stationary) 시계열 데이터입니다.

ADF 검정, KPSS 검정 결과:

비정상 시계열 데이터 (KPSS 불만족)

ADF 검정, KPSS 검정 결과:

정상 시계열 데이터

4. 추가 연구

변수 추가 분석 및 모델에 적절한 변환 필요

구분	기온	풍향	풍속	강수량	습도	태양광 발전량	풍력발전량
데이터 타입	float	float	float	float	float	float	float



구분	기온	풍향	풍속	강수량	습도	태양광 발전량	풍력발전량
데이터 타입	float	float	float	float	float	float	float

2023.11.22

제안서_팀_빵빵이들



팀 구성

빵빵	201902743 - 이의진
빵일?	201802798 - 이정현
빵이	201903156 - 정경서
빵삼	202103464 - 최서여